HEALTH EFFECT DATABASES:
MAYOR'S TASK FORCE ON THE HEALTH EFFECTS OF AIR POLLUTION

INSTITUTE FOR HEALTH POLICY

The University of Texas School of Public Health at Houston

TABLE OF CONTENTS

TYPES OF DATABASES

Hospital Discharge

National Hospital Discharge Survey (NHDS)

- The NHDS is Provided by the National Center for Health Statistics
- Available for 1996-2003
- Datasets available online
- Includes "Information on characteristics of inpatients discharged from non-Federal short-stay hospitals in the United States. The NHDS collects data from a sample of approximately 270,000 inpatient records acquired from a national sample of about 500 hospitals. Only hospitals with an average length of stay of fewer than 30 days for all patients, general hospitals, or children's general hospitals are included in the survey."
- Variables include age, sex, race, ethnicity, marital status, and expected sources of payment. Administrative items such as admission and discharge dates (which allow calculation of length of stay), as well as discharge status are also included. Medical information about patients includes diagnoses and procedures coded to the *International Classification of Diseases, 9th Revision, Clinical Modification* (**ICD-9-CM**). Since 1977, patient zip code, expected source of payment, and dates of surgery have also been collected**. (The medical record number, date of birth, and patient zip code are confidential information and are not available to the public)**.
- Data are classified by location in one of the four geographic regions of the United States corresponding to those used by the U.S. Bureau of the Census.

Texas Health Care Information Collection Center for Health Statistics

- The Texas Health Care Information Collection Center for Health Statistics collects information from hospitals and health maintenance organizations.
- **Hospital discharge data** come in two forms:
    - **Texas Public Use Data File (PUDF)**
    - **Data for Research**
- Texas Public Use Data File (PUDF). The PUDF contains data on discharges from Texas hospitals. Data are available by quarter for 1999 through 2003. The Texas Public Use Data File (PUDF) is available for purchase (**see Attachment A**). The PUDF contains data on hospital inpatient discharges and includes:
    - Patient demographic data - including age group, sex, race and ethnicity
    - Geographic data - including the patient's zip code
    - Diagnoses and procedures
    - Primary and secondary sources of payment – including detailed charges
- Data for Research. "Research files" are Hospital discharge data not released in the PUDF and can be requested for research purposes by individuals and organizations, including universities and government agencies. The data elements required for the proposed research and a description of the intended use of the data are reviewed by the THCIC. The approved application is then submitted to the DSHS Institutional Review Board, acting as the Scientific Review Panel (SRP), for final approval.
    - Data Elements Available for Research – **See Attachment B** for variables included in dataset. Patient address is included in Research File.
    - Research Data Request Form – **See Attachment C**
    - Research Data Use Agreement

> o **Research Data Request Process** – **See Attachment D**. Request is reviewed by DSHS IRB for merit and human subject protection

Emergency Room/Ambulatory Surgery

The National Hospital Ambulatory Medical Care Survey (NHAMCS)

- The National Hospital Ambulatory Medical Care Survey (NHAMCS) is designed to collect data on the utilization and provision of ambulatory care services in hospital emergency and outpatient departments.
- Annual data collection began in 1992.
- Data are obtained on demographic characteristics of patients, expected source(s) of payment, patients' complaints, physicians' diagnoses, diagnostic/screening services, procedures, medication therapy, disposition, types of health care professionals seen, causes of injury where applicable, and certain characteristics of the hospital, such as type of ownership (**see Attachment E** for data collection form). **See Attachment F** for list of variables for outpatient department visits and emergency department visits.
- Please note geographic variables are U.S. Census geographic region and Metropolitan Statistical Area (MSA) or non-MSA location of visit.
- Files can be downloaded at the National Center for Health Statistics, Ambulatory Health Care Data website. **See Attachment G** for more information.

Other Hospital Information

- Both pediatric and adult asthma hospitalizations have been compiled and reported by county by the Texas Department of State Health Services.

Mortality

National Center for Health Statistics

- Mortality Data from the National Vital Statistics System – National Vital Statistics System provides mortality data on 50 states, two cities and five territories. Information is compiled by death certificates provided by each state and medical information is coded with the International Classification of Diseases. Geographic classification includes Metropolitan Statistical Areas (MSAs), Primary Metropolitan Statistical Areas (PMSAs), and New England County Metropolitan Areas (NECMAs). There are also 20 Consolidated Metropolitan Statistical Areas (CMSAs) which are made up of PMSAs. State, county, and city level data.
- Data Set Characteristics **–** Each record on the annual data set contains underlying cause (coded using ICD-10), demographic, and geographic detail and two multiple cause-of-death fields coded using ICD-10. Geographical data include the state, city, county of residence and Primary Metropolitan Statistical Areas and Metropolitan Statistical Areas, FIPS (**see Attachment H**).

CDC Wonder

- [Compressed Mortality File, Underlying Cause-of-Death](#) – The Compressed Mortality database contains mortality and population counts for all U.S. counties for the years 1979 to 2002. Counts and rates of death can be obtained by underlying cause of death, state, county, age, race, sex, and year (**see Attachment I**).
- [National Occupational Mortality Surveillance](#) (NOMS) – Interactive database with state level data, 1960-1994 (**see Attachment J**).
- [NIOSH Mortality](#) – The NIOSH mortality data set has counts and rates of death for underlying cause of death for any of the 92 standard groups of causes by a single by variable: age, race, gender, year, state and country of residence, or group of cause, 1960-1994 (**see Attachment J**).
- [MMWR](#) – Mortality data by Region and Major Cities from 1996 to current; includes Corpus Christi, Dallas, El Paso, Ft. Worth, San Antonio and Houston)

Texas Health Data

- [Death Data](#) Interactive database at county level **(see attachment K)**
- [Texas Health Data--Mortality Maps](#) (Texas Department of State Health Services, 1990 to 1998 [ICD-9; **county** level data). Interactive site for developing maps and retrieve data to look at deaths by gender and age as well as by disease or cause of death and year. Data at county level.

Cancer

American Cancer Society

- [Statistics](#) – The ACS tracks cancer occurrence, including the number of deaths, cases, and how long people survive after diagnosis. ACS also tracks data regarding behaviors that influence the risk of developing cancer and the use of screening tests. Data are available on state level.

American College of Surgeons

- [National Cancer Data Base (NCDB)](#) – NCDB and specifically benchmark reports provide data in state level for variable such as age, cancer primary site, sex, age, race, and hospital type. Table and graphs can be generated.

Centers for Disease Control and Prevention

- [United States Cancer Statistics: 2001 Incidence and Mortality](#) – Data at state level by cancer type, sex, ethnicity, age.

National Cancer Institute

- Cancer Mortality Maps and Graphs – This web site provides interactive maps, graphs, text, tables and figures showing geographic patterns and time trends of cancer death rates for the time period 1950-1994 for more than 40 cancers. Data on county and SEA (state economic area) level (**see Attachment L**)
    - o Cancer Survivor Prevalence Statistics
    - o Geographic Information Systems – This is not a primary health database. Cancer researchers may use this site to see some examples of how their colleagues apply GIS technology in their work. In addition, this site contains many links to GIS resources and information on the Web, which may be helpful to researchers interested in incorporating GIS in their work.
    - o Cancer State Profile – Interactive maps and interactive databases. Data are available in county level by other variables such as cancer site, sex, age, and race.
    - o SEER (Surveillance, Epidemiology, and End Results)
        - ▪ SEER Incidence and U.S. Mortality Statistics – The Cancer Query Systems (CANQUES) are data retrieval applications that provide access to cancer statistics stored in online databases. These systems do not perform calculations; they display reports using databases of statistics generated by other software. Many of these statistics are provided in the SEER Cancer Statistics Review or other SEER Statistical Publications. The Cancer Query Systems have a Web-based interface that allows you to retrieve the statistics relevant to your research that you may not have found in a published format. Through this interface, you can:
            - • select the type of statistic and stratification variables used in the report
            - • select the format of the report by choosing from various table and graph formats
            - • extract the statistics, in a delimited format, for further analyses in other software.
        - ▪ SEER provides Cancer Statistics in a variety of formats, including the Cancer Query Systems listed above, SEER Statistical Publications, and the Fast Stats system.
        - ▪ US Mortality Data **(See Attachment L)**
        - ▪ Surveillance, Epidemiology, and End Results (SEER) Program, Trends in Incidence and Mortality by Race and Age – Fast Stat is a useful tool in providing data on state level about cancers incidence and mortality.
        - ▪ SEER-Medicare Linked Database – The SEER-Medicare data reflect the linkage of two large population-based sources of data that provide detailed information about elderly persons with cancer. The data come from the Surveillance, Epidemiology and End Results (SEER) program of cancer registries that collect clinical, demographic and cause of death information for persons with cancer and the Medicare claims for covered health care services from the time of a person's Medicare eligibility until death (**see Attachment M**).

National Center for Health Statistics

- Cancer mortality rates and counts of deaths for the United States and for individual states are based on causes of death reported by the certifying physicians on death certificates filed in state vital statistics offices. The mortality information is processed and consolidated into a national database by the NCHS of the Centers for Disease Control and Prevention (Healthy Women: State Trends in Health and Mortality).

North American Association of Central Cancer Registries

- Cancer incidence rates and counts of cases for individual states are obtained from NAACCR, which sets standards for population-based registries and works toward compatibility of methods. In 1999, NAACCR provided cancer incidence statistics for 42 central cancer registries for cases diagnosed during 1991-1995 (Cancer in North America Publications).

Texas Cancer Registry

- The Texas Cancer Registry (**see Attachment N**) provides cancer incidence and mortality statistical tables, and information on how to request special data sets or data linkage. Statewide cancer incidence data are currently available from the TCR for Texas resident cancer cases diagnosed from 1995-2002. Interactive database at county level. Further confidential information provided upon request.

Birth Defects

Texas Department of State Health Services (TDSHS)

- Texas Health Data, Birth Defects Data – Interactive database with birth Defect Data at county level (**see Attachment O**).

Cardiovascular Diseases

Centers for Disease Control and Prevention

- Heart Disease and Stroke Maps (US Centers for Disease Control & Prevention; 1991 to 2001; **state** and **county** level data)

USERS OF HEALTH EFFECT DATABASES FOR ENVIRONMENTAL STUDIES

City of Houston Environmental Public Health Tracking Program (EPHT)

As part of Public Health Environmental Tracking Initiative, the Houston Department of Health and Human Services (HDHHS) is developing a local environmental public health tracking system that builds on existing local data systems and integrates with the Texas Department of Health's chronic disease, injury, and birth defects surveillance systems; the Texas Natural Resource Conservation Commission's monitoring data; and the U.S. Environmental Protection Agency's tier-two data to provide a local response, surveillance, and research resource to

mitigate the effects of environmental exposures on the health of Houstonians. The existing databases being used are:

- Texas Cancer Registry
- Texas Birth Defects Registry
- Texas Health Care Information Collection Hospital Discharge Data
- BRFSS – (optional indoor air module used)

Other cities are involved in the PHET, and the indicators used are seen in **Attachment P**.

HELIX – Atlanta

The Centers for Disease Control and Prevention (CDC) is coordinating HELIX-Atlanta to provide information regarding the five-county Metropolitan Atlanta Area (Clayton, Cobb, DeKalb, Fulton, and Gwinett) via a network of integrated environmental monitoring and public health data systems so that all sectors can take action to prevent and control environmentally related health effects. The HELIX-Atlanta Network is a tool to access interoperable information systems with optional information technology linkage functionality driven by scientific rationale. HELIX-Atlanta is a collaborative effort with local, state, federal, and academic partners. The HELIX-Atlanta Partners identified the following HELIX-Atlanta initial focus areas: childhood lead poisoning, short-latency cancers, developmental disabilities, birth defects, vital records, respiratory health, age of housing, remote sensing, and environmental monitoring.

Public Health Air Surveillance Evaluation (PHASE)

The purpose of PHASE is to develop, evaluate, and demonstrate the advantages and limitations of different methods of generating air quality characterization surveillance data that can be systematically and routinely available to link with public health surveillance data as part of the EPHT Network. The evaluation portion of this project is scheduled for completion by September 2005. At the present, no results have been disseminated (**see Attachment Q**). Participants are:

- CDC's National Center for Environmental Health
- EPA's National Exposure Research Laboratory
- EPA's Office of Air Quality, Standards, and Planning
- Wisconsin Division of Public Health
- New York State Department of Health
- Maine Bureau of Health